

Minutes from the Database and Related Utilities meeting 11/1/02 12:30 – 1:30 EST

Attending:

Dave Davis, Joanne Bogart, Pat Nolan, Robert Schaefer, and Karl Young.

Minutes

After a brief “hello” and check out of VRVS connectivity, we discussed the work that needs to be done, the assignment of specific tasks, and the schedule for completing the tools.

First, Bob asked if anyone had any comments about the tasks listed in the link from the agenda page. Karl mentioned that Richard duBois looked at the list and suggested that the priorities could be adjusted. In particular, he suggested that looking at hardware items like disk types should not be high on the priority list, presumably because hardware can change so rapidly. Bob agreed with this assessment, but mentioned that we didn’t want to ignore hardware concerns either.

We then proceeded through the list of tasks:

- We discussed whether RDBMS indexing had really been tested adequately. Also using a RDBMS is our backup plan if we have too much trouble with the cluster software solution, so it is worthwhile to continue investigating RDBMS systems. Pat offered to try exercising an indexed RDBMS, since he already has the HTM code working.
- Karl and Bob offered to look at cluster software solutions. For the cluster software, Karl mentioned that it would be useful to have some machines to test out different cluster software. For test hardware, Bob and Dave agreed to look at test hardware at Goddard, and Karl agreed to assess what hardware would be available at SLAC.
- To pick parallelization code – Bob offered to look up information about the MPI and PVM, etc. Karl mentioned that it seemed the support for PVM had disappeared and MPI had emerged as the standard.
- SQL. A lengthy discussion emerged about whether SQL needed to be implemented for a cluster based D1. The consensus was that for D1, we probably do not need to have an SQL interface. We should have a standard set of queries that need to be supported by D1 which are a small subset of the possibilities offered by SQL. Then if we need to switch to an RDBMS, the U1 tool can rephrase those queries into SQL for the RDBMS. Bob agreed to ask other groups to help make a list of necessary D1/U1 queries.
- SCSI/ATA, etc. disk types. Pat and Karl had already looked into this issue and volunteered to collect their notes and make them available to us.
- Mirror sites. Bob said that, currently, we just need to identify contacts at mirror locations that we can talk to. Dave agreed to find out who at ASI we need to contact. Traudl has offered to talk to the right people at Saclay. Karl mentioned that the people at Saclay require that the data be in an Oracle database. Joanne

commented that this would then not be a mirror since they are specifying their own implementation. Traudl is away at CERN for two weeks and can hopefully comment at or before our next VRVS.

- D2 pixelization scheme. Since the pixelization requirement on the database has been not well defined, Pat was coerced into looking at the needs of the exposure calculator to see if we need to pixelize the data, and if so, what kind of pixelization.
- D6 – other catalogs. Bob commented that perhaps we need to wait for the other science tool groups to tell us what catalogs will be needed. Dave suggested that we support and interface to NED, which would give us access to many other catalogs.
- Communication between analysis tools and database extractors. Bob commented that a “standard search” of 6 month duration yields about 100 Mb of data, so transfer of data to the client needs to handle this speed bottleneck. Bob agreed to talk to HEASARC about what methods they use for handling remote queries and what the security folks recommend for accepting remote queries.

Bob pointed out that according to the schedule drawn up by Seth, we are supposed to have finished the D2 prototype by Jan1, 2003, and the D1 prototype by Apr. 1, 2003 (no joke!), so we need to stay focused.

We agreed to have VRVS meetings every other week. Our next meeting will be Nov. 15, 2002.

Bob encouraged people to use the database Wiki server (the link can be found from the main database and related utilities web page).

Action items:

Pat

- test DBMS using an indexed search
- look at exposure calculator needs for pixelization in D2.
- Collect notes on SCSI/ATA disks and make it available to group (Wiki page would be nice).

Bob

- look at possible cluster test hardware and see if we can get more development accounts on LHEA's Beowulf.
- Collect information about MPI,PVM, etc.
- Ask other groups what types of D1 queries are needed for analysis.
- Ask other groups to tell us what catalogs they want to use.
- Ask HEASARC about what procedures/protocols they would recommend for doing remote client-side queries to our servers.

Karl

- See if cluster test hardware is available at SLAC.

- Collect notes on SCSI/ATA disks and make it available to group (Wiki page would be nice).
- Look at cluster “management” software.

Dave

- see if we can get cluster test hardware (2-3 machines?) for studying cluster operation
- Find out who is our D! mirror contact at ASI.